DOCUMENT RESUME

ED 180 364 HE 012 101

AUTHOR Posey, Ellen I.: Sylvan, Donna L.
TITLE Concerns in the Selection of an Appropriate Data Source: A Critical Aspect of Institutional Research.
PUB DATE Oct 79
NOTE 16p.: Paper presented at the Annual Meeting of the Southern Association for Institutional Research (Orlando, FL, October 1979)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Data Analysis: *Data Collection: Decision Making: Higher Education: *Information Sources: *Institutional Research: Researchers: *Research Methodology: *Research Problems

ABSTRACT

The selection of appropriate data sources for institutional research is discussed. It is suggested that the provision of timely, accurate information for administrative decision-making depends on the researcher's ability to determine which data are relevant to the problem. This is of particular concern for institutional researchers because they frequently do not create the data for their analyses and must be especially sensitive in their selection of a data source. An example of a typical institutional descriptor, average class size, is used to clarify the issues involved in the selection of an appropriate data source. It is suggested that within the same institution different data sources can be available for measures with the same label and that varying definitions generate the data sources. Special attention is focused on the definitional aspect of the data source as a critical point in the assessment of the appropriateness of the source for a given problem. Limitations of the sources are explored. It is recommended that institutional researchers be cognizant of the limitations imposed by the definition of the data, and make the consumers of their analyses aware of these limitations. (Author/SF)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED180364

CONCERNS IN THE SELECTION OF AN APPROPRIATE DATA SOURCE: A CRITICAL ASPECT OF INSTITUTIONAL RESEARCH

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

Ellen I. Posey

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

by

Ellen I. Posey

and

Donna L. Sylvan
Office of Institutional Research
Georgia State University
Atlanta, Georgia 30303

Paper Presented to the Annual Meeting
of the Southern Association for Institutional Research
Orlando, Florida
October, 1979

HE 012 101

CONCERNS IN THE SELECTION OF AN APPROPRIATE DATA SOURCE: A CRITICAL ASPECT OF INSTITUTIONAL RESEARCH

Donna L. Sylvan
and
Ellen I. Posey

Office of Institutional Planning
Georgia State University
Atlanta, Georgia 30303
404-658-2570

A vital function of institutional research is the provision of timely, accurate information for administrative decision making. A critical aspect of this function is the ability to assess the actual question asked, to determine and process the necessary data, and to provide decision makers with information which they can utilize. This paper focuses on one step in this process: the determination of an appropriate data source. This emphasis is necessary because data which are not relevant to the problem provide no information. Furthermore, this problem is of particular concern for institutional researchers because they frequently do not create the data for their analyses. They, therefore, must be especially sensitive in their selection of a data source, selecting that source which is defined in a manner which can yield data relevant to the question at hand.

An example using a typical institutional descriptor, average class size, serves to clarify the issues involved in the selection of an appropriate data source. It is shown that within the same institution different data sources can be available for measures with the same label and that varying definitions generate the data sources. Special attention is focused on the definitional aspect of the data source as a critical point in the assessment of the appropriateness of the source for a given problem. Limitations of the sources are explored.

Institutional researchers must be sensitive to the problems inherent in selecting an appropriate data source for analysis. In particular they must be cognizant of the limitations imposed by the definition of the data, and they must make the consumers of their analyses aware of these limitations.

Concerns in the Selection of an Appropriate Data Source:
A Critical Aspect of Institutional Research

One vital function of institutional research is to provide timely, accurate information for administrative decision making. A critical aspect of this function often depends upon a three-step process involving the ability to assess the question asked, to determine and process the necessary data and to provide decision makers with information which they can utilize effectively. This paper focuses primarily on the issues regarding the assessment of appropriate data sources for an institutional research function. Emphasis on these issues is necessary because data which are not relevant to the problem provide little useful information. These issues are of particular concern for institutional researchers because they frequently do not create the data for their analyses. In order to select an appropriate data source capable of yielding information relevant to a given question, the question must first be adequately examined thereby enabling both administrators and institutional researchers to define and understand the problem at hand. The researcher then incorporates this definition in the selection of an appropriate data source. This paper also focuses on other issues relative to the selection of data sources such as the institution's administrative data perspective and the extent of fragmentation prevalent in the institution's data storage systems.

An example using a typical institutional descriptor, average class size, serves to clarify some of the issues involved in the selection of an

appropriate data source. Within the same institution, several different data sources exist as measures of class size which carry the same label, although these measures often have differing definitions and are perceived differently by the various operational components within the institution.

The Research Process

A vital role often expected of institutional researchers is to provide the institution's administration with timely decision-making information. The research aspect of this role involves: 1) assessment of the questions, 2) determination of appropriate data source(s), and 3) the end result of relevant, timely and accurate information. The importance of data definition is present at each step of this research process. Interpretation of data always depends upon the assumption that answers to questions asked are based upon working definitions of the concepts and constructs involved. Selltiz, Jahoda, Deutsch and Cook (1959) strongly emphasize the importance of definitions:

> The greater the distance between one's concepts or constructs, and the empirical facts to which they are intended to refer, the greater the possibility of their being misunderstood or carelessly used, and the greater the care that must be given to defining them (p.41).

For researchers to assess questions accurately, they must make certain that they define problems in the same way that administrators have conceptualized them. If the problem is not accurately defined, an inadequate and inappropriate data source may be selected, thereby failing to yield results which furnish the requested information. In other words,

care in the definition of the problem enhances the likelihood that researchers and administrators are "talking about the same side of the coin."

Data definitions are also important in the selection of an appropriate data source. Once consensus has been reached on the definition of the problem, the researcher is faced with the task of finding a data base which is relevant to the problem as defined. The research process once again necessitates definitional consensus. The definition of the problem must be congruent with the data definitions used to construct items in the selected data source.

The providing of decision-making information is an important step in the research process. If the key concepts are not consistently defined and understood by all parties, the data items will not be relevant to the problem and in all likelihood will not provide reliable and valid summary information. The existence or gathering of individual data items does not constitute or guarantee that useful information can be derived. Information is a "set of elements needed to make a current decision or to serve as the premise for a future decision and raw data always need to be aggregated, transformed, interpreted and appropriately conveyed in order to qualify as information" (Balderston, 1974:48).

Administrative Perspective and Policy

Although institutional research offices may differ from institution to institution, most are expected, at some point, to respond effectively to administ ative requests for information. The range of understanding which administrators have of the research process and managerial policies are also important variables to consider.

Higher education is no longer in a position where the absence or implementation of policy can be ignored. Many institutions are now being held accountable for certain issues and problems which have existed for a long time (Bacchetti, 1977). In essence, many institutions are now facing problems which require a concern for policies related to institutional data definitions, operational procedure, and institutional concepts and purposes.

The lack of well-defined institutional policies is often reflected in the development and confused condition of an institution's internal data systems. The absence of policy and holistic perspective concerning the intended and prospective uses of data bases have sometimes resulted in the development and tiresome maintenance of fragmented data storage systems. In the absence of a comprehensive institution-wide data management policy many college and university data systems have developed haphazardly according to the changing needs of the unit managers responsible for non-standardized operational procedures.

Where policy is found to exist it is often inadequate either in terms of focus (mechanical versus definitional) or range of utilization (operational units versus secondary users). The focus on operational mechanics for the exclusive convenience of a primary operational unit often renders the primary data meaningless to all other potential users but that particular primary operational unit. The data bases are, in the absence of an effective holistic policy, therefore not designed and maintained with the various needs of potential and secondary users in mind.

Data Selection Concerns

The providing of information for administrative decision making and policy formulation is often a key element in the development of decisions and/or policies which can influence greatly the functioning of the institution, e.g., response to a legal challenge. It is important to realize that information must be created by a process which includes defining, collecting, evaluating, interrelating, analyzing, and interpreting accurate, reliable and timely primary data (Gubasta and Kaufman, 1977). The providing of such information is sometimes the responsibility and always the concern of the institutional researcher.

An administrative request for information may require the institutional researcher to gather data directly in the form of questionnaires and surveys or to utilize existing institutional data. In either case, to respond to requests effectively, the researcher must be aware of the array of extant internal and external data sources and sources of potential information via direct data collection. In the course of their research, institutional researchers must be able to respond to requests from audiences with various viewpoints and levels of expertise and provide analyses which encompass broad viewpoints. In other words, institutional researchers must be able to assess present and future needs in order to respond effectively to various administrative requests. To accomplish this end, the research analyst must incorporate a global perspective and have access to and knowledge of available primary data about the institution (Saunders, 1979).

Unfortunately, primary operational data systems do not always provide the most appropriate source for institutional research and informational needs. Control and use of the primary data systems are are often lively issues in an institution and often pose problems for the institutional researcher. In addition, the research analyst often has to question the integrity of the data as a source of reliable information. The existence of primary data systems, therefore, does not insure that they are the most appropriate and accessible source to draw upon. Yet, the problem often arises that the existent operational data sources are the only sources available.

Saunders (1979) illuminates several problems inherent in using institutional data systems: 1) The systems are often unintegrated and fragmented. 2) Each system is a separate activity and each new system added or modified utilizes the latest innovations. However, all systems usually are not updated to utilize the innovations, thus constituting an inter-system lag. 3) Data definitions between systems will vary because they are developed at different points in time. 4) Data definitions between systems will vary and some may overlap but the result is still incomplete and fragmented data sources.

The characteristic of "unintegrated systems" causes several problems for the institutional researcher. First, because of the fragmentation, the researcher may be required to combine incompatible data from several sources. Then, he frequently runs into trouble because of inconsistent definitions between operational units. Furthermore, the

efforts to consolidate incompatible data constitutes a strain upon the researcher's resources. The reliability of the data is an additional issue for the researcher. Even though operational units usually edit and understand their data, such operational units have a tendency to edit their data in light of criteria relevant to their level of functioning. This practice seldom results in appropriate and reliable data needed for institutional research purposes.

Example of Data Source Selection: Average Class Size Reports

Some of the data-related problems which are frequently encountered in institutional research are represented in the following exercise. Administrators often express an interest in information about changes in average class size. To respond to this interest, the researcher needs first to explore the issue or problem presented, and then the questions asked, and the information desired to determine an appropriate data source.

Attempts to clarify the administrative interest in average class size did not yield a specific statement of the issue in question. The general theme appeared to be an interest in understanding differences in average class size over time. In this instance average class size was interpreted as an indicator of the number of students served per class/instructor. It was anticipated that useful institutional information should be available. Given the vague statement of the problem, the decision was made to explore several institutional data sources as a preliminary step in determining an appropriate data source to identify factors which might influence average class size.

Of the institutional data sources considered, two appeared promising. To assess the adequacy of these sources with regard to the current task five aspects of the sources were considered: the primary source of the data, the method of average class size computation, accessibility, time span, detail level, and purpose of the report. It should be noted that while both sources originated within the same office and were similarly labeled "average class size," their content differed. These differences appeared to reflect inconsistency of purposes and data selection procedures as a result of no existing policy defining a common indicator of average class size for the institution.

Quarterly reports of average class size were seen as one potential source. In these reports the primary data used to determine average class size were class count reports as produced by the Registrar's office following the last date for registration and course changes. Manual adjustments were made in these reports to ensure that 1) cross-listed courses were counted under only one academic department, 2) some types of courses were omitted, and 3) team-taught courses were counted as one class. After these manual exclusions and exceptions, the average class size figure was manually computed by dividing the number of students by the number of sections. The reports displayed information by department for each college but information by level of instruction, e.g., upper or lower division, was not available consistently. These quarterly average class size reports were generated to provide admininstrators with a measure which generally represented the average number of students in a section or class.

Another potential source of average class size information was provided within an official, institutional document. This information was aggregated from the teaching assignment file also prepared by the Registrar. In these reports sections taught by more than one person had all teachers counted with each receiving credit for a portion of the students in the class. An attempt was made to include all classes so that courses excluded from the quarterly reports previously discussed were included in the determination of class size in this document. The method for computing average class size was the same as that used in the quarterly report, dividing the number of students by the number of sections. This source yielded an average class size measure developed to reflect average teaching loads.

The quarterly reports were chosen as a source of analysis in the preliminary study for the following reasons. First, average class size information was defined in a manner which was consistent with the intent of the study. In addition, although some subjective judgment was used in determining the actual measure, this information was considered to be consistent in its development. Third, this source also provided the largest amount of historical data and was readily available for the researcher's use.

Discussion

As the example points out, the factors researchers must consider in the selection of an appropriate data source are numerous. Although the selected source may not constitute the best possible choice, it may be the most efficient and feasible of the sources currently at his disposal.

Even though the institutional researcher realizes the limitations of the data sources available to him, it is difficult for him as a secondary user to make adequate impact for data definitional consistency. An ideal situation would be for institutional researchers and data systems personnel to work together so that the needs of both the researcher and the operational entities be met. Idealistically, this cooperation would result in definitions that are consistent, integrated and flexible in terms of the potential for institution-wide use.

There have been many solutions proposed to rectify this problem ranging from the establishment of basic data management policies to the fostering of institution-wide cooperation in data systems and usage. Gubasta and Kaufman (1977) have outlined a viable solution:

1) There should be a constant monitoring of the sources of operational data as a check on the utility of such data for planning and management purposes
2) Administrators should identify the issues and concerns they wish to have addressed before data are collected and analyzed
3) The persons or offices where data are generated should be tapped as a source of expertise and insights into the quality and meaning of the data provided
4) Sufficient time should be allowed for data analyses before reporting
5) Procedures for information collection and update should be thoroughly documented so that tasks can be performed by

clerical staff under supervision. This would assure continuity of efforts and comparability of data

6) Information should be disseminated as widely as organizational policy will permit. This would help to negate the need for colleges and departments to maintain separate records or hire analytical studies staff of their own

7) Efforts to increase the number of organizational units covered should be made (p. 41)

All of these factors illustrate the fact that if institutional researchers are to cope with the demands for timely administrative information, they must somehow influence the initial design and application of data storage systems. A first step is to seek ways to make administrative personnel aware of the importance of data definitions and that these definitions are the vital linkage between administrative policy, operational procedure, data storage systems, and the ultimate production of administrative information. Researchers should advocate an administrative global perspective of the problem and encourage administrators to participate actively. Administrators themselves must be able to define and interpret issues important to the institution before any effective policies can be conceptualized or established. Adequate information which would clarify these issues is not apt to be economically available if comprehensive institution-wide data management policies are nonexistent. Primary operational data will not emerge as decision-making information until the institutional aims and concepts can be defined clearly and operationally.

Conclusion

An analysis of the research process shows the importance of data definition consensus among institutional researchers, administrators and data generators/collectors. Because the researcher is often forced to use existing data bases, he, as a secondary user, has no power over the initial design of the data base. However, the institutional researcher's global perspective of the institution's operation, informational needs, and research procedure, justify his participation in the conceptual and design stage of many operational systems. In order to impress the need for a total perspective, the researcher should assume an "advocacy role" with the administration advocating the need for a holistic approach to the establishment of policies regarding data definitions and an unfragmented and accessible data system. This role would also require the researcher to advocate a collective perspective to data source problems and solutions. All parties would benefit from collective efforts and collaboration in terms of efficient use of institutional resources and the accuracy and timeliness of response to administrative needs and requests.

References

Bacchetti, R. F. In praise of good questions. New Directions for Institutional Research, 1977, 15, (Autumn), 11-18.

Balderston, F. E. The design and uses of information systems. New Directions for Institutional Research, 1974, 1, (Spring), 47-66.

Gubasta, J. L. & Kaufman,N. Developing information for academic management: An alternative to computer-based systems. Journal of Higher Education, 1977, 58, (4), 401-411.

Saunders, L. E. Dealing with information systems: The institutional researcher's problems and prospects. The AIR Professional File, 1979, 2, (Summer), 1-4.

Selltiz, C., Jahoda, M., Deutsch, M.,& Cook, S. W. Research Methods in Social Relations. New York: Holt, Rinehart and Winston, 1959.